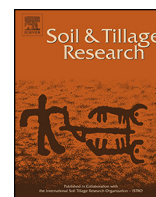


This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



On-line measurement of soil properties without direct spectral response in near infrared spectral range

Omar Marín-González^{a,*}, Boyan Kuang^b, Mohammed Z. Quraishi^b, Miguel Ángel Munóz-García^a, Abdul Mounem Mouazen^b

^a Rural Engineering Department, Electrotechnical Section, EUITAgrícola, UPM, Avenida Complutense, s/n, 28040 Madrid, Spain

^b Environmental Science and Technology Department, Cranfield University, Bedfordshire MK43 0AL, United Kingdom

ARTICLE INFO

Article history:

Received 5 December 2012

Received in revised form 17 April 2013

Accepted 24 April 2013

Keywords:

On-line measurement

Fertility

Vis–NIR spectroscopy

Accuracy

Independent validation

ABSTRACT

So far, the majority of reports on on-line measurement considered soil properties with direct spectral responses in near infrared spectroscopy (NIRS). This work reports on the results of on-line measurement of soil properties with indirect spectral responses, e.g. pH, cation exchange capacity (CEC), exchangeable calcium (Ca_{ex}) and exchangeable magnesium (Mg_{ex}) in one field in Bedfordshire in the UK. The on-line sensor consisted of a subsoiler coupled with an AgroSpec mobile, fibre type, visible and near infrared (vis–NIR) spectrophotometer (tec5 Technology for Spectroscopy, Germany), with a measurement range 305–2200 nm to acquire soil spectra in diffuse reflectance mode. General calibration models for the studied soil properties were developed with a partial least squares regression (PLSR) with one-leave-out cross validation, using spectra measured under non-mobile laboratory conditions of 160 soil samples collected from different fields in four farms in Europe, namely, Czech Republic, Denmark, Netherland and UK. A group of 25 samples independent from the calibration set was used as independent validation set. Higher accuracy was obtained for laboratory scanning as compared to on-line scanning of the 25 independent samples. The prediction accuracy for the laboratory and on-line measurements was classified as excellent/very good for pH (RPD = 2.69 and 2.14 and r^2 = 0.86 and 0.78, respectively), and moderately good for CEC (RPD = 1.77 and 1.61 and r^2 = 0.68 and 0.62, respectively) and Mg_{ex} (RPD = 1.72 and 1.49 and r^2 = 0.66 and 0.67, respectively). For Ca_{ex} , very good accuracy was calculated for laboratory method (RPD = 2.19 and r^2 = 0.86), as compared to the poor accuracy reported for the on-line method (RPD = 1.30 and r^2 = 0.61). The ability of collecting large number of data points per field area (about 12,800 point per 21 ha) and the simultaneous analysis of several soil properties without direct spectral response in the NIR range at relatively high operational speed and appreciable accuracy, encourage the recommendation of the on-line measurement system for site specific fertilisation.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

During the last decade visible near infrared spectroscopy (vis–NIRS) is being increasingly used to detect soil properties, with variable accuracy depending on several factors (Kuang et al., 2012). The vis–NIRS is a simple, non-destructive and rapid technique, needs no sample preparation for field applications, and can be used for the laboratory, in situ (Viscarra Rossel et al., 2006a), and on-line measurements (Mouazen et al., 2005). Furthermore, in some cases accuracy obtained is high and very similar to that of conventional procedures (Viscarra Rossel et al., 2001). The technique allows the assessment of primary properties with direct spectral responses, which are directly affected by combinations and overtones of

fundamental vibrations for organic functional groups and water, particle size, and surface properties (Chang et al., 2001). In a review paper, Stenberg et al. (2010) defined organic carbon (OC), total carbon (TC), moisture content (MC) and clay minerals to have direct spectral response in the NIR spectroscopy. In this context, it is possible to detect clay type and content, MC, OC and TC with high accuracy (Volkan et al., 2010). Other soil properties without direct spectral absorption features in the vis–NIR range (secondary properties) can be also measured with good to moderate accuracy due to co-variation with one or more primary properties (Stenberg et al., 2010). For both soil categories, colour plays an important role in enhancing measurement accuracy. Good results have been reported for cation exchange capacity (CEC), pH, extractable calcium (Ca_{ex}) and extractable magnesium (Mg_{ex}) under non-mobile laboratory and in situ conditions but underperformed those for properties with direct spectral responses in the NIR range, e.g. OC, TC, MC and clay content (Kuang et al., 2012). According to the

* Corresponding author. Tel.: +34 689616130; fax: +34 915449983.

E-mail addresses: o.marin@upm.es, o.marin.gon@gmail.com (O. Marín-González).

literature, sodium (Na_{ex}) and potassium (K_{ex}) are among the most difficult properties to be measured with the NIR spectroscopy (Malley et al., 1999; Chang et al., 2001; Zornoza et al., 2008; Pirie et al., 2005; Dunn et al., 2002; Shepherd & Walsh, 2002; Islam et al., 2003; Volkan et al., 2010; Mouazen et al., 2006). For the same soil property, laboratory vis–NIR methods achieved higher accuracy as compared to measurement under field soil conditions, particularly with on-line vis–NIR sensors (Stenberg et al., 2010; Kuang et al., 2012).

The prediction accuracy achieved so far with the on-line vis–NIRS sensors available today (Shibusawa et al., 2001; Mouazen and Ramon, 2006; Christy, 2008) might be sufficient for many applications in precision agriculture, since spatial and temporal variation of soil properties is large relative to the precision of measurement (Shepherd and Walsh, 2002). However, similar to laboratory and in situ measurements, Kuang et al. (2012) concluded that the best accuracy is achieved for soil properties with direct spectral response, which is probably the reason why researchers using on-line vis–NIRS sensors have focused mainly on soil properties with direct spectral responses (Shonk et al., 1991; Mouazen et al., 2005; Bricklemeyer and Brown, 2010; Munoz and Kravchenko, 2011; Knadel et al., 2011). Mouazen et al. (2007) showed potential success for the on-line vis–NIRS measurement of extractable and available P and pH, without proving the accuracy to be of quantitative meaning. Although Mouazen et al. (2007) demonstrated spatial similarities between measured and on-line predicted pH, no robust conclusions on accuracy of measurement could be drawn. Later Mouazen et al. (2009) conducted on-line measurement of available P with remarkable accuracy (RPD = 1.42; $r^2 = 0.62$). To our knowledge none of the previous studies has reported on the on-line measurement of CEC, Ca_{ex} and Mg_{ex} . In addition, no comparison was made between laboratory and on-line measurement of the named soil properties.

The aim of this paper is to evaluate the performance and accuracy of on-line measurement of soil properties without direct spectral responses in the NIR spectroscopy range, namely, CEC, pH, Ca_{ex} and Mg_{ex} . It also aims to compare the prediction accuracy of these properties based on soil spectra collected under on-line measurement conditions with those collected under laboratory non-mobile conditions. Calibration models developed for several farms in Europe will be used to validate the on-line measurement of these properties in one selected field in the UK.

2. Materials and methods

2.1. Soil samples

A total of 140 soil samples were used to develop general calibration models for the prediction of pH, CEC, Ca_{ex} and Mg_{ex} . These soil samples were collected from different fields in six farms in four different European countries (Fig. 1 and Table 1). A total of 25 samples were collected from one field in Mespil Medlov, A.S. farm (Czech Republic), 20 samples from two fields in Bramstrup

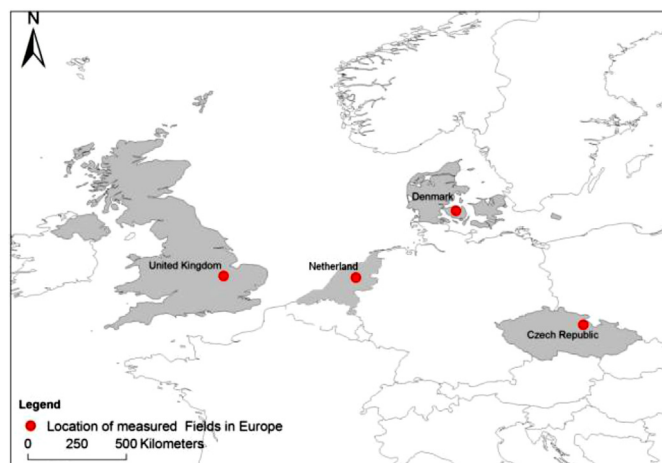


Fig. 1. Locations of the fields where soil samples were collected to establish general calibration models for soil pH, cation exchange capacity (CEC), extractable calcium (Ca_{ex}) and magnesium (Mg_{ex}) were collected.

Estate farm (Denmark), 23 samples from one field in Wageningen University (the Netherlands), 25 samples from one field in Ely farm (Cambridgeshire, UK), 17 samples from five different fields in Silsoe experimental farm (Bedfordshire, UK) and 30 samples from fields 2 and 3 in Duck End farm (Bedfordshire, the UK) (Table 2). Bulk samples were collected from the upper soil layer (0–30 cm) in the spring of 2008 in the Mespil Medlov, A.S. farm and in the spring of 2009 in Bramstrup Estate farm (Kuang and Mouazen, 2011). For the remaining farms, soil samples were collected from the bottom of 15 cm deep trenches opened by a subsoiler during the on-line vis–NIR measurement in the autumn of 2010 in Wageningen University, summer of 2009 in Silsoe experimental farm, spring of 2011 in Ely farm, and summer of 2011 in Duck End farm (Mouazen and Ramon, 2006). These fields correspond to a large diversity of soil textures, crops and landscapes (Table 1).

In order to validate the calibration models developed, on-line measurement was carried out in field 1 in Duck End Farm (Table 2). During this measurement a total of 45 samples were collected for model development and validation. About 44% of these samples (20 sample) were added to the general calibration set (140 + 20 = 160 samples) to develop the general calibration models of the selected four properties, whereas the remaining 25 samples were used as validation set for the validation of the general calibration models based on non-mobile (laboratory) and on-line measured spectra.

Approximately 200 g of each soil sample was kept deep frozen (–18 °C) until testing. Before analysis, each sample was defrosted, carefully mixed and divided into two portions, with one used for chemical analysis, and the other used for optical measurement. Samples were stored in plastic bags at 4 °C during the analysis to avoid losing humidity.

Table 1
Information about the fields in the farms in Mespil Medlov, A.S. in Czech Republic, Bramstrup Estate in Denmark, Wageningen University in the Netherlands and Silsoe, Ely and Duck End in the UK, where soil samples were collected.

Farm	Area, ha	Crop	Texture type	Sand, %	Silt, %	Clay, %
Czech Republic	3	Wheat	Silt clay loam	4.86	70.58	24.56
Denmark	2	Wheat	Sandy loam	68.57	21.96	9.48
Netherlands	1	Maize	Silty clay loam	6.17	66.22	27.61
UK (Silsoe)	16	Wheat	Clay loam	40.11	27.38	32.51
UK (Duck End)	37.1	Winter Barley	Clay loam	42.46	29.45	28.09
UK (Ely)	1.5	Lettuce	Silty clay	2.25	48.03	49.72

Table 2

Soil samples used to establish and validate general calibration models for measurement of soil pH, cation exchange capacity (CEC), extractable calcium (Ca_{ex}), and magnesium (Mg_{ex}) (Kuang and Mouazen, 2011).

Country	Number of samples	Number of Fields	Sampling time
Czech Republic	25	1	2008
Denmark	20	2	2009
Netherland	23	1	2010
UK (Silsoe)	17	5	2009
UK (Duck End, field 1)	45	1	2011
UK (Duck End, field 2)	10	1	2011
UK (Duck End, field 3)	20	1	2011
UK (Ely)	25	1	2011

2.2. Laboratory measurement of soil properties

Laboratory analyses of pH, CEC and base saturation, were performed at the soil laboratory of the National Soil Resource Institute, School of Applied Sciences (NR-SAS) of Cranfield University (Bedfordshire, the UK) using their standard procedures. Soil pH was measured in a 1:5 soil:H₂O suspension following the BS ISO 10390 (2005) for the determination of pH. In order to determine the CEC and exchangeable cations (Na_{ex} , K_{ex} , Ca_{ex} , Mg_{ex}), the air-dried soil samples were first saturated with respect to barium by adding 30 ml of barium chloride solution (Reagent Production Unit (RPU) 10), after which 30 ml excess of 0.02 mol/l magnesium sulphate (RPU 11) was added. This makes all barium present in the solution as well as adsorbed in exchangeable sites to precipitate in the form of highly insoluble barium sulphate. As a result, the exchangeable sites are occupied by magnesium. The surplus of magnesium was determined by atomic absorption following the NR-SAS SOP 42/Version 1, based on the BS 7755 Section 3.12 (1996), for the determination of the potential cation exchange capacity and exchangeable cations using barium chloride buffered at pH = 8.1, which is identical to ISO 13536 (1995) (Table 3). Base saturation was calculated as the equivalent sum of major base cations (Ca_{ex} , Mg_{ex} , K_{ex} and Na_{ex}) percentage of CEC (Chodak et al., 2004).

2.3. Optical measurement

2.3.1. Optical measurement in laboratory

For each soil sample a certain amount of soil was mixed up in a glass bowl. Stones and plant residues were removed at this point. Three small cups of 1 cm deep and 3.6 cm in diameter were filled up with the same soil sample. The surface of the samples was softly pressed and smoothed down with a spatula, simulating the effect

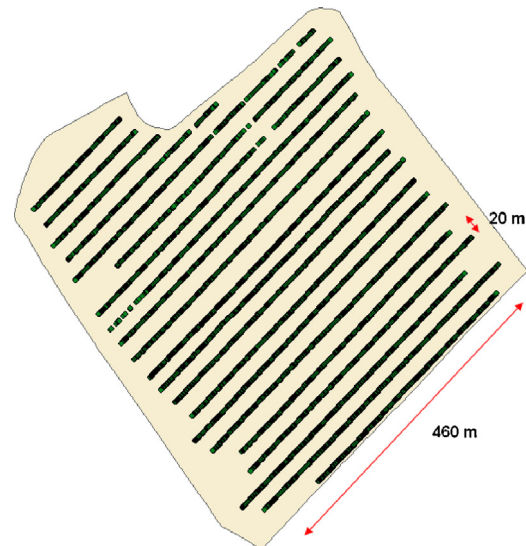


Fig. 2. On-line measured lines in the field 1 in the Duck End Farm (Bedfordshire, the UK).

of the subsoiler smoothing of soil beneath the chisel during on-line measurement, which increases the signal to noise ratio (Mouazen et al., 2005).

An AgroSpec mobile, fibre type, vis-NIR spectrophotometer (Tec5 Technology for Spectroscopy, Germany) with a measurement range of 305–2200 nm was used to measure soil spectra in diffuse reflectance mode. A 100% ceramic was used as the white reference, which was scanned once every 30 min. Optical scanning was conducted on non-treated, fresh soil samples to simulate field measurement conditions. A total of ten scans were performed for each of the three bowels prepared for each soil sample. The resulted thirty spectra were averaged into one spectrum for each sample (Kuang and Mouazen, 2011). This averaged spectrum was used for spectra pre-treatment and model development. All calibration (160) and validation (25) samples were scanned in the laboratory.

2.3.2. On-line vis-NIR measurement

The on-line spectra were collected along parallel measurement lines in the field 1 in the Duck End Farm in the UK (Fig. 2). Detailed information about this field is shown in Table 4. The same vis-NIR spectrophotometer used for laboratory measurement (AgroSpec) was used for the on-line measurement. The on-line sensor developed by Mouazen (2006) was used to carry out the on-line measurement. The subsoiler makes a trench in the soil, whose

Table 3

Laboratory methods of soil analyses.

Soil property	n^a	Technique	References
pH _w	185	1:5 soil:H ₂ O extract	BS ISO 10390 (2005)
CEC cmolc/kg	171	BaCl ₂ buffered at pH = 8.1	BS 7755 Section 3.12 (1996)
Exchangeable Na cmolc/kg	171	BaCl ₂ buffered at pH = 8.1	BS 7755 Section 3.12 (1996)
Exchangeable K cmolc/kg	171	BaCl ₂ buffered at pH = 8.1	BS 7755 Section 3.12 (1996)
Exchangeable Ca cmolc/kg	171	BaCl ₂ buffered at pH = 8.1	BS 7755 Section 3.12 (1996)
Exchangeable Mg cmolc/kg	171	BaCl ₂ buffered at pH = 8.1	BS 7755 Section 3.12 (1996)

^a Number of samples used for the analysis.

Table 4

Information about the field 1 in the Duck End farm, Bedfordshire (UK), where on-line soil measurement took place in summer, 2011.

Field	Area, ha	Crop	Number of Samples Nr	Texture	Sand, %	Silt, %	Clay, %
Field 1	21	Winter barley	20 + 25	Clay loam	41.92	30.57	27.51

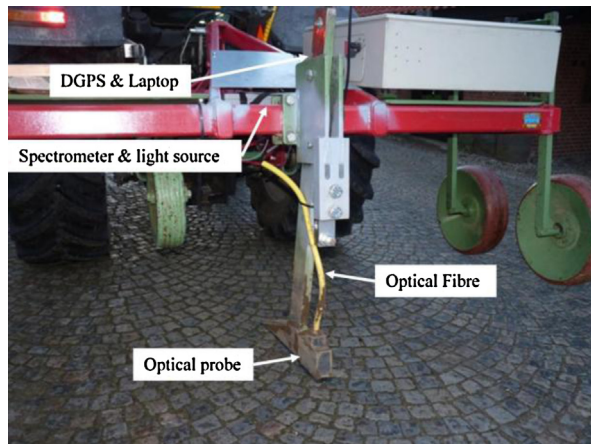


Fig. 3. The on-line visible and near infrared (vis-NIR) spectroscopy-based sensor, attached to the tractor for on-line measurement (Mouazen, 2006).

bottom is smoothened by the subsoiler itself, due to the downwards vertical forces. The optical probe protected in a steel lens holder was appended to the backside of the subsoiler chisel to measure soil spectra in diffuse reflectance mode from the smoothened bottom of the 15 cm deep trench created by the subsoiler chisel. The subsoiler retrofitted with the optical probe was attached to a frame, which was mounted onto the three point linkage of a tractor (Fig. 3). The spectrophotometer was IP 64 protected for harsh working environments. A differential global positioning system (DGPS) (EZ-Guide 250, Trimble, USA) was used to record the position of on-line measured spectra with sub-metre accuracy. A Panasonic semi-rugged laptop was used for data logging and communication. The spectrophotometer, laptop and DGPS were powered by the tractor battery.

2.4. Pre-processing of spectra

Several spectra pre-processing were tested and the best performing one was kept. The selection criteria of any pre-processing were the largest coefficient of multiple determination (r^2) and residual prediction deviation (RPD), which is the ratio of standard deviation (S.D.) of the prediction data set to root mean square error of prediction (RMSEP) and the smallest RMSEP. Spectra pre-processing and establishment of calibration models of different properties were done by Unscrambler 7.8 software (Camo Inc., Oslo, Norway). To remove noise at edges of spectra, soil spectrum was first abridged to 400–2100 nm for each sample. After noise was removed, spectra were reduced by averaging three successive wavelengths in the visible range (400–1000 nm) and six successive wavelengths for the NIR region (1000–2100 nm). Averaging over wavelengths was used to decrease the number of wavelengths and to smooth the spectrum (Nicola et al., 2007). Maximum normalisation was followed, which is typically used to get all data to approximately the same scale, or to get a more even distribution of the variances and the average values. This method attempted to remove the effects of scattering by linearising each spectrum to some 'ideal' spectrum of the sample, which, in practice, corresponds to the average spectrum (Nicola et al., 2007). The maximum normalisation 'polarizes' the spectra. The peaks of all spectra with positive values are scaled to +1, while spectra with negative values are scaled to -1. The peaks of these spectra were scaled to +1, since all soil spectra in this study had positive values (Mouazen et al., 2005). The maximum normalisation was selected because it provided better results for all properties considered compared with other pre-processing tested. Spectra were subsequently subjected to Savitzky–Golay first derivation (Martens and

Naes, 1989). This transformation procedure generally intensifies the absorption characteristics indicative of soils properties, and diminishes variation among spectra (Volkan et al., 2010). This method enabled the computation of the first or higher-order derivatives, including a smoothing factor, which determines how many adjacent variables should be used to estimate the polynomial approximation used for derivatives. A second-order polynomial approximation was selected with a 2:2 smoothing factor. A 2:2 Savitzky–Golay smoothing was carried out after the first derivative to remove random noise from spectra (Kuang and Mouazen, 2011).

2.5. Establishment of calibration models

The pre-processed spectra and the results of laboratory chemical analyses were used to develop calibration models for all studied properties. Before partial least squares regression (PLSR) analysis, the 160 calibration samples (Tables 1 and 2) were divided into calibration (85% of samples) and prediction sets (15% of samples). Further validation of developed models was done using the 25 samples collected in the field 1 in Duck End farm, which was designated as independent validation set.

To develop calibration models, especial attention should be paid in the selection of calibration and validation sets. The distribution and size of the calibration data set must cover the entire range of concentration of a soil property examined. To build a robust calibration model, validation of model developed has to be carried out using validation samples, which were not used for model development. This means that validation samples of the independent validation set (25 samples collected from the field 1 in the Duck End farm) should not be used for the development of calibration models in cross-validation, avoiding this way any influence in the prediction capacity of the model selected and the consequent overestimation (Brown et al., 2005). The same samples of calibration and validation were selected for each property. A previous threshing of samples was executed to get the wider range of values and spectra variation in the calibration set. Calibration and validation samples were then selected randomly. Performing this way a calibration data set was set to cover the whole range of concentration for the different soil properties examined.

The most used multivariate methods to develop calibration models are based on linear regressions. Mainly, stepwise multiple linear regression (SMLR), principal component regression (PCR), and PLSR (Stenberg et al., 2010) are used. The PLSR with leave-one-out cross-validation was carried out using Unscrambler 7.8 software (Camo Inc., Oslo, Norway) to generate the calibration models relating soil independent variables (wavelengths) of the diffuse reflectance spectra to each soil parameter. PLSR performs particularly well, compared with other multivariate statistical methods, when there is a high dimensional correlation between variables, which is the case for soil spectral data (Volkan et al., 2010). Also, PLSR is favoured because it requires fewer components to explain the variance in the response, due to the relation that this method establishes between response and predictor variables, and its results are more interpretable (Stenberg et al., 2010).

The number of latent variables for a model was determined by examining a plot of leave-one-out cross-validation residual variance against the number of latent variables obtained from PLSR. The latent variable of the first minimum value of residual variance was selected (Brown et al., 2005). The residual sample variance and predicted vs. measured plots were assessed for outliers determination after running the PLSR. Outliers may be induced by typing errors, file transfer, interface errors, sensor malfunctions and fouling, poor sensor calibration, bad sampling or sample presentation, etc. (Nicola et al., 2007). Samples located individually far from the zero line of residual variance together

Table 5

Statistics of laboratory analyses of soil samples of calibration (85% of samples) and prediction (15% of samples) sets.

Property	Samples	Calibration set				Prediction set			
		Min	Max	Mean	SD ^a	Min	Max	Mean	SD ^a
pH	160	4.87	8.43	7.28	0.89	5.36	8.19	7.08	0.79
CEC (cmolc/kg)	146	7.30	21.10	16.97	3.27	10.70	20.40	17.23	2.66
Ca _{ex} (cmolc/kg) ^b	122	9.91	135.13	40.05	21.53	17.33	72.68	34.23	13.43
Mg _{ex} (cmolc/kg) ^c	122	0.43	7.69	1.51	0.95	0.57	3.93	1.41	0.79

^a Standard deviation.^b Ca_{ex} after excluding Ely farm samples.^c Mg_{ex} after excluding Ely farm samples.

with a far position from the trend line in the predicted vs. measured plot were considered as outliers and excluded from the analysis. A maximum of 5% of the entire samples was accepted as the maximum number of outliers to be removed (Kuang and Mouazen, 2011).

2.6. Performance assessment of calibration models

Prediction accuracy of a PLSR model was determined by the RMSEP, r^2 and RPD. The criteria adopted for RPD classification (Viscarra Rossel et al., 2006b) was as follows: RPD < 1.0 indicates very poor model/predictions and their use is not recommended; RPD between 1.0 and 1.4 indicates poor model/predictions where only high and low values are distinguishable; RPD between 1.4 and 1.8 indicates fair or moderately good model/predictions which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicates good model/predictions where quantitative predictions are possible; RPD between 2.0 and 2.5 indicates very good, quantitative model/predictions, and RPD > 2.5 indicates excellent model/predictions. This classification system was adopted in this study.

3. Results and discussion

For robust modelling of the vis-NIR spectral data, the selection of calibration set should be carefully done so that to be representative of the samples used for validation. The range for sample concentrations in the prediction (Table 5) and independent validation (Table 6) sets is smaller and within the corresponding range for the calibration set (Table 5) for all properties. This confirms that variation in the prediction and independent validation sets are accounted for in the calibration set.

Since samples of Ely field were of excessively high exchangeable Ca_{ex} and Mg_{ex} values, the 25 samples from Ely field were disposed during the development of the calibration models for these two properties, as these might negatively affect the prediction accuracy of the corresponding models.

3.1. Evaluation of general calibration models

Examining the results of the prediction set (Table 7), reveals that best results are achieved for Mg_{ex} ($r^2 = 0.88$; RPD = 2.55) and pH ($r^2 = 0.86$; RPD = 2.37). Less accurate estimations were obtained for CEC ($r^2 = 0.72$; RPD = 1.70) and Ca_{ex} ($r^2 = 0.76$; RPD = 1.87). According to the classification based on RPD proposed by Viscarra Rossel et al. (2006b), the prediction accuracy in the prediction set is excellent and very good for Mg_{ex} and pH, respectively, good for Ca_{ex} and moderately good for CEC. These results are in line with those found in the literature. Comparing with other studies reported in the literature under laboratory measurement condition (Shepherd and Walsh, 2002; Cohen et al., 2005; Mouazen et al., 2006; Viscarra Rossel and Behrens, 2010), the model performance for pH in the prediction set is among the best models ($r^2 = 0.50$ –

0.97; RMSEP = 0.04–1.43; RPD = 0.57–2.39). Similar conclusion can be drawn for the prediction of Mg_{ex}, with overall results reported in the literature for r^2 of 0.53–0.91; RMSEP of 0.03–38.36 cmol kg⁻¹ and RPD of 0.48–2.54 (Cozzolino and Moron, 2003; Groenigen et al., 2003; Udelhoven et al., 2003; Wetterlind et al., 2010). Almost the same results as those achieved in this study were reported by Dunn et al. (2002), when analysing pH for the top soil of 0–10 cm ($r^2 = 0.80$, RPD = 2.3) and Mg_{ex} ($r^2 = 0.85$, RPD = 2.7). The models for the other two properties are less accurate, as compared with other values reported in the literature for Ca_{ex} ($r^2 = 0.07$ –0.95; RMSEP = 0.66–52.90 cmol kg⁻¹;

Table 6

Statistics of laboratory analyses of the independent validation set samples collected in the field 1 in Duck End farm (UK).

Property	Samples	Min	Max	Mean	SD ^a
pH	25	5.34	8.41	6.66	0.77
CEC (cmolc/kg)	25	10.40	18.00	14.62	1.74
Ca _{ex} (cmolc/kg) ^b	25	5.91	44.91	24.16	9.69
Mg _{ex} (cmolc/kg) ^c	25	0.57	2.08	1.08	0.49

^a Standard deviation.^b Ca_{ex} excluding Ely farm samples.^c Mg_{ex} excluding Ely farm samples.**Table 7**Performance of the general calibration models of soil pH, cation exchange capacity (CEC), extractable calcium (Ca_{ex}) and magnesium (Mg_{ex}), validated with prediction set, laboratory scanned and on-line scanned independent validation set (25 sample) in the field 1 in the Duck End farm in the UK.

Property	r^{2a}	SD	RMSEP ^b	RPD ^c
Full cross validation				
pH	0.79	0.86	0.40	2.16
CEC	0.58	2.72	1.77	1.54
Ca _{ex}	0.89	18.38	22.05	0.83
Mg _{ex}	0.69	4.22	0.38	2.44
Prediction				
pH	0.86	0.79	0.33	2.37
CEC	0.72	2.50	1.47	1.70
Ca _{ex}	0.76	13.77	7.34	1.87
Mg _{ex}	0.88	0.78	0.31	2.55
Laboratory independent validation				
pH	0.86	0.77	0.28	2.69
CEC	0.68	1.74	0.99	1.77
Ca _{ex}	0.86	9.70	4.43	2.19
Mg _{ex}	0.66	0.50	0.29	1.72
On-line independent validation				
pH	0.78	0.76	0.36	2.14
CEC	0.62	1.56	0.97	1.61
Ca _{ex}	0.61	9.29	7.11	1.30
Mg _{ex}	0.67	0.51	0.34	1.49

^a Coefficient of determination.^b Root mean square error of prediction in cmol kg⁻¹ for CEC, Ca_{ex} and Mg_{ex}.^c Residual prediction deviation (standard deviation/root mean square error of prediction).

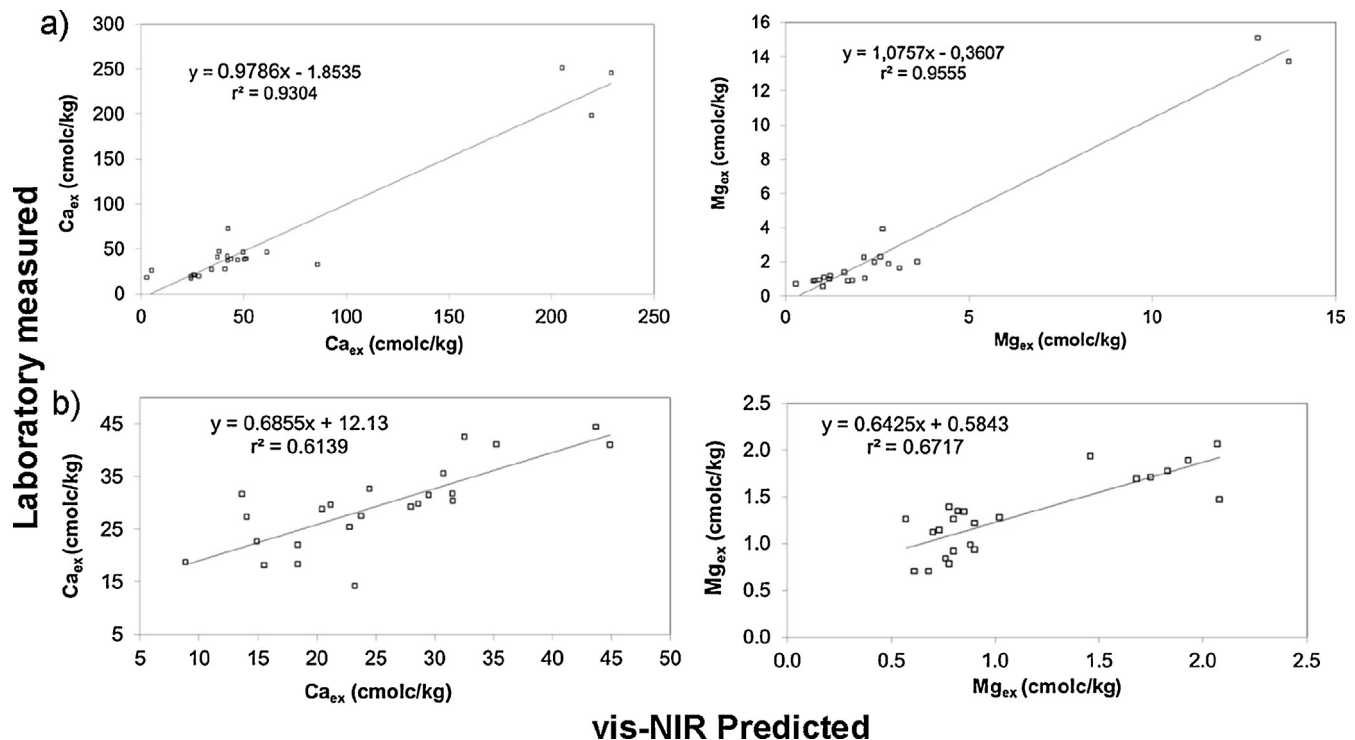


Fig. 4. Scatter plots between laboratory and on-line visible and near infrared (vis-NIR) predicted values for extractable calcium (Ca_{ex}) and (Mg_{ex}) before removing Ely field samples (a) and after removing Ely field samples (b).

RPD = 0.60–2.75) (e.g. Cozzolino and Moron, 2003; Cohen et al., 2005; Mouazen et al., 2006; Zornoza et al., 2008) and CEC ($r^2 = 0.13$ – 0.90 ; RMSEP = 1.22 – 10.43 cmol kg⁻¹; RPD = 0.60–2.7) (e.g. Ben-Dor and Banin, 1995; Chang et al., 2001; Mouazen et al., 2006; Awiti et al., 2008). Islam et al. (2003), reported similar results for CEC ($r^2 = 0.64$, RPD = 1.6), using a separate validation set. However, these accuracies can still be considered as good and useful for quantitative predictions.

3.2. Comparison of laboratory and on-line measurement accuracy

In order to compare the performance of the general calibration models for the prediction of studied soil properties between laboratory and on-line scanned spectra, the independent validation set (25 samples) collected from the field 1 in Duck End farm was used. Generally, smaller accuracies are observed when using soil spectra collected under on-line measurement conditions, as compared to that for spectra collected under laboratory non-mobile scanning conditions (Table 7). Results confirm that the predictions of the laboratory and on-line measurements were classified as excellent/very good for pH (RPD = 2.69 and 2.14 and $r^2 = 0.86$ and 0.78 , respectively), and moderately good for CEC (RPD = 1.77 and 1.61 and $r^2 = 0.68$ and 0.62 , respectively) and Mg_{ex} (RPD = 1.72 and 1.49 and $r^2 = 0.66$ and 0.67 , respectively). For Ca_{ex}, very good accuracy was calculated for the laboratory method (RPD = 2.19 and $r^2 = 0.86$), as compared to the poor accuracy for the on-line method (RPD = 1.30 and $r^2 = 0.61$).

There is little literature about the on-line prediction of secondary soil properties e.g. properties without direct spectral responses in the NIR spectroscopy. Soil pH was the most successfully measured property with on-line vis-NIR sensors, with all reports show less accurate on-line predictions than the ones achieved in the current study ($r^2 = 0.78$) with r^2 values of 0.62 (Christy, 2008) and 0.61 (Shibusawa et al., 2001). Only comparison between measured and predicted pH maps was provided by

Mouazen et al. (2007), showing similar trends of spatial distribution. The low prediction accuracy obtained in this study for on-line Ca_{ex} might be attributed to the uneven distribution of sample concentration over the entire concentration range. In spite of the wide range of concentration, the majority of samples have low values for Ca_{ex} while there is a small group of samples with very high values (Fig. 4). Thus, there is a gap of values between both groups, hindering the creation of a robust calibration model. Similar comment can be made for laboratory scanned spectra (data not shown). Similar trend is observed for Mg_{ex} with a lower impact on the accuracy for both laboratory non-mobile and on-line measurement. To reduce this effect samples from Ely field (UK), with the highest values for Ca_{ex} and Mg_{ex}, were removed during the development of calibration models, which led to the results shown in Table 7 and Fig. 4. In the future, it is recommended to consider new samples with Ca_{ex} and Mg_{ex} values covering the gap in the concentration range for the development of more robust calibration models. The low accuracy in predictions of the Ca_{ex} model could also be related with the extraction method selected. This method might be inappropriate for Ca_{ex} detection in predominantly calcareous samples.

3.3. Analysis of error

3.3.1. Histogram of error

The histogram of normal distribution plots of error was calculated by subtracting predicted from measured values for each property using 25 samples of the independent validation set (Table 8). It can be clearly observed that the mean, variance and SD values are much larger with on-line scanned than with laboratory scanned spectra. This is a clear sign to support the conclusion that accuracy decreases during on-line measurement due to the ambient conditions (e.g. noise, vibration, stones, plant roots, and a spectrum position with a corresponding soil sample) (Mouazen et al., 2007; Stenberg et al., 2010). However, under laboratory

scanning conditions all these ambient conditions affecting accuracy are eliminated.

3.3.1.1. Histograms of error for pH. The histogram plot of normal error distribution for pH prediction using laboratory scanned spectra is normally distributed around 0 (Fig. 5 a), with slight skewness (0.392) towards the positive side, indicating slight

overestimation. However, for on-line prediction, a smaller skewness (0.024) can be observed (Table 8). Around 76% and 64% of laboratory and on-line predictions, respectively, are with a smaller error than 0.3 in absolute values, which is smaller than 12% of the normal pH range for agricultural soils (4–9) (USDA, 1998). Furthermore, the range of prediction error is larger for on-line prediction (–1.003 to 1.122), as compared to laboratory prediction.

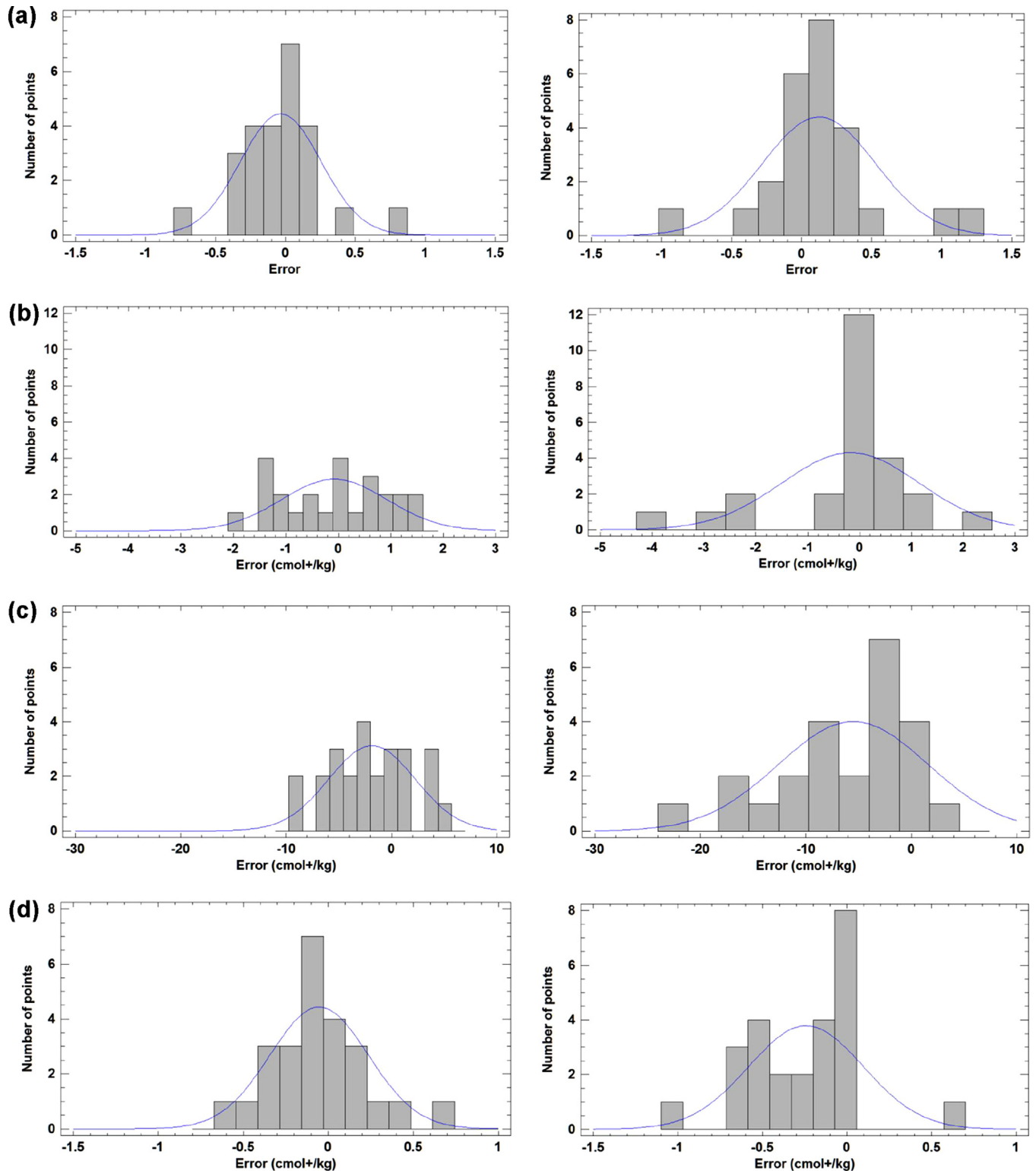


Fig. 5. Histogram of normal distribution of error for laboratory (left) and on-line (right) predictions of soil pH (a), cation exchange capacity (CEC) (b), extractable calcium (Ca_{ex}) (c) and magnesium (Mg_{ex}) (d) in the independent validation set of the field 1 in the Duck End farm in the UK.

Table 8

Summary of statistics for the histogram plots of error for on-line and laboratory scanned spectra.

Elements	Skewness	Kurtosis	Mean	Variance	SD	Error range
Laboratory independent validation						
pH	0.392	1.649	−0.034	0.083	0.289	−0.674 to 0.762
CEC	−0.004	−1.098	−0.079	1.004	1.002	−1.840 to 1.611
Ca _{ex}	−0.010	−0.738	−1.858	16.859	4.106	−9.349 to 5.022
Mg _{ex}	0.491	1.489	−0.053	0.084	0.289	−0.655 to 0.733
On-line independent validation						
pH	0.023	2.949	0.124	0.164	0.405	−1.003 to 1.122
CEC	−1.201	2.324	−0.176	1.761	1.327	−3.908 to 2.483
Ca _{ex}	−0.523	0.439	−5.511	50.620	7.115	−21.911 to 9.025
Mg _{ex}	0.016	0.744	−0.242	0.115	0.338	−1.010 to 0.608

The smaller range of prediction error in addition to the larger portion of small error than 0.3 (76%) of laboratory scanned spectra suggests the better performance of the general pH model for prediction based on laboratory than on-line scanned spectra.

3.3.1.2. Histograms of error for CEC. Similar to pH, the histogram plots of CEC error for both laboratory and on-line predictions show normal distributed around 0 (Fig. 5b), with skewness values of −0.004 and −1.201, respectively (Table 8). Around 64% and 76% of laboratory and on-line predictions, respectively, are with smaller error than 1.1 cmol+/kg, which are smaller than 4.4% of the normal CEC range for agricultural soils CEC ranges from 0 cmol+/kg for sandy soils to about 50 cmol+/kg for clay soils (Mengel and Kirkby, 1982). This surprisingly shows a bigger proportion of error around 0 for on-line prediction, as compared the laboratory prediction. However, the range of error was larger for on-line than for laboratory prediction (Table 8). This might be attributed to difficulties associated with matching position of a soil sample collected for laboratory chemical analysis and the corresponding spectrum collected during on-line measurement (Mouazen et al., 2007).

3.3.1.3. Histograms of error for Ca_{ex}. The normal distribution plot of error for Ca_{ex} prediction using laboratory scanned spectra (Fig. 5c) is clearly skewed towards the negative side of the plot. This skewness is larger for on-line (skewness = −0.523) than for laboratory (skewness = −0.010) predictions (Fig. 5c, Table 8). About 80% and 52% of laboratory and on-line predictions, respectively, are with smaller error than 5.5 cmol+/kg in absolute values, which are smaller than about 9% of the Ca_{ex} range present in the samples. This confirms that laboratory prediction of Ca_{ex} is more accurate than the on-line prediction, which is supported by a larger range of error for the on-line, as compared to the laboratory prediction (Table 8).

3.3.1.4. Histograms of error for Mg_{ex}. The normal distribution plot of error of Mg_{ex} of laboratory predictions presents a rather normal distribution of error around 0. However, this is clearly skewed towards the negative range when on-line spectra are used (Fig. 5d, Table 8). About 56% and 48% of laboratory and on-line predictions, respectively, are with smaller error than 0.17 cmol+/kg in absolute values, which are a smaller error than about 4.7% of the Mg_{ex} range present in the samples. This indicates that as for pH and Ca_{ex}, the prediction of Mg_{ex} is more accurate and stable when laboratory scanned spectra are used, as compared to on-line measurement. This did not prove clearly for CEC, although the trend is similar to the other three properties.

The above discussion about histogram plots of error and accuracy analysis confirms that calibration models developed with laboratory scanned spectra perform better when validated with laboratory measured spectra, as compared to on-line measured

spectra. However, in comparison with laboratory chemical and vis-NIR analyses, the on-line measurement of soil properties enables the collection of high number of data points (around 12,800 readings for 21 ha of the field 1), with average of around 2 points per metre travel distance. This large amount of data allows the spatial interpolation to estimates values for un-sampled points in the field. Then, the possibility to predict several soil properties from the same spectrum, the opportunity to create maps from this large amount of information and its utilisation for site specific land management within the field, suggest the on-line measurement of soil properties without direct spectral responses in the NIR range as a valuable measuring technique.

4. Conclusions

This paper reports on the performance of a vis-NIR spectroscopy-based on-line sensor for the prediction of soil properties without direct spectral response in the NIR range, namely, pH, CEC, Ca_{ex} and Mg_{ex}. It also compares prediction accuracy of these properties between on-line and non-mobile laboratory scanning. The results obtained in this study allow the following conclusions to be drawn:

1. The on-line measurement system enabled the simultaneous measurement of several soil properties without direct spectral responses in the NIR spectroscopy across a field.
2. General calibration models, developed with samples collected from fields in different European counties are a successful procedure for the calibration of the on-line vis-NIR sensor for the prediction of pH, CEC and Ca_{ex} and Mg_{ex}.
3. Higher accuracy was obtained for predictions using laboratory scanned, as compared to on-line scanned spectra. The laboratory and on-line predictions were classified as excellent/very good for pH (RPD = 2.69 and 2.14 and r^2 = 0.86 and 0.78, respectively), and moderately good for CEC (RPD = 1.77 and 1.61 and r^2 = 0.68 and 0.62, respectively) and Mg_{ex} (RPD = 1.72 and 1.49 and r^2 = 0.66 and 0.67, respectively). For Ca_{ex}, very good accuracy was calculated for the laboratory method (RPD = 2.19 and r^2 = 0.86), as compared to the poor accuracy for the on-line method (RPD = 1.30 and r^2 = 0.61).
4. The histogram plots of error proved the general calibration models developed with laboratory scanned spectra to perform better when used to predict the studied soil properties using laboratory scanned spectra, as compared to on-line scanned spectra.
5. The ability of continuous data gathering with the on-line soil sensor at a relatively high operational speed (about 3 km/h) with very good to moderate accuracy obtained for some of the properties investigated (e.g. pH, CEC and Mg_{ex}), suggest the recommendation of the on-line soil sensor for site specific fertilisation.

Further research is needed to upgrade the calibration models of pH, CEC, Ca_{ex} and Mg_{ex} , developed in this study using samples collected from a larger number of fields and countries with even distribution of concentrations along the entire concentration range encountered in agricultural soils. This is recommended to improve the prediction accuracy and robustness of models developed for studied soil properties.

References

- Awiti, A.O., Walsh, M.G., Shepherd, K.D., Kinyamario, J., 2008. Soil condition classification using infrared spectroscopy: a proposition for assessment of soil condition along a tropical forest-cropland chronosequence. *Geoderma* 143, 73–84.
- Ben-Dor, E., Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal* 59, 364–372.
- Brickley, R.S., Brown, D.J., 2010. On-the-go VisNIR: potential and limitations for mapping soil clay and organic carbon. *Computers and Electronics in Agriculture* 70, 209–216.
- British Standard, BS ISO 10390. (2005). Determination of pH.
- British Standard, BS 7755 Section 3.12. (1996). Determination of the potential cation exchange capacity and exchangeable cations using barium chloride buffered at pH = 8.1.
- Brown, D.J., Brickley, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of vis-NIR soil C prediction in Montana. *Geoderma* 129, 251–267.
- Chang, C., Laird, D.A., Mausbach, M.J., Hurburgh Jr., C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal* 65, 480–490.
- Chodak, M., Khanna, P., Horvath, B., Beese, F., 2004. Near infrared spectroscopy for determination of total and exchangeable cations in geologically heterogeneous forest soils. *Journal of Near Infrared Spectroscopy* 12, 315–324.
- Christy, C.D., 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Computers and Electronics in Agriculture* 61, 10–19.
- Cohen, M.J., Prenger, J.P., DeBusk, W.F., 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. *Journal of Environmental Quality* 34, 1422–1434.
- Cozzolino, D., Moron, A., 2003. The potential of near-infrared reflectance spectroscopy to analyze soil chemical and physical characteristics. *Journal of Agricultural Science* 140, 65–71.
- Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of near-infrared reflectance for soil analysis—a case study from the Riverine Plain of south-eastern Australia. *Australian Journal of Experimental Agriculture* 42, 607–614.
- Groenigen, J.W., Muters, C.S., Horwath, W.R., Kessel, C., 2003. NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. *Plant Soil* 250, 155–165.
- Islam, K., Singh, B., McBratney, A.B., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research* 41, 1101–1114.
- Knadel, M., Thomsen, A., Greve, M.H., 2011. Multisensor on-the-go mapping of soil organic carbon content. *Soil Science Society of America Journal* 75, 1799–1806.
- Kuang, B., Mouazen, A.M., 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *European Journal of Soil Science* 62 (4), 629–636.
- Kuang, B., Mahmood, H.S., Quraishi, Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, E.J., 2012. Sensing soil properties in the laboratory, in situ, and on-line: a review. In: Sparks, D. (Ed.), *Advances in Agronomy*, vol. 114. Academic Press, Agron, UK, pp. 155–224.
- Malley, D.F., Yesmin, L., Wray, D., Edwards, S., 1999. Application of near-infrared spectroscopy in analysis of soil mineral nutrients. *Communications in Soil Science and Plants Analysis* 30, 999–1012.
- Martens, H., Naes, T., 1989. *Multivariate Calibration*, second ed. John Wiley and Sons Ltd., Chichester, UK, pp. 419 pp.
- Mengel, K., Kirkby, E.A., 1982. *Principles of Plant Nutrition*, third ed. International Potash Institute Bern, Switzerland, pp. 655 pp.
- Mouazen, A.M., De Baerdemaeker, J., Ramon, H., 2005. Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. *Soil & Tillage Research* 80, 171–183.
- Mouazen, A.M. (2006). Soil Survey Device. International publication published under the patent cooperation treaty (PCT). World Intellectual Property Organization, International Bureau. International Publication Number: WO2006/015463; PCT/BE2005/000129; IPC: G01N21/00; G01N21/00.
- Mouazen, A.M., Ramon, H., 2006. Development of on-line measurement system of bulk density based on on-line measured draught, depth and soil moisture content. *Soil and Tillage Research* 86, 218–229.
- Mouazen, A.M., De Baerdemaeker, J., Ramon, H., 2006. Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 14 (3), 189–199.
- Mouazen, A.M., Maleki, M.R., De Baerdemaeker, J., Ramon, H., 2007. On-line measurement of some selected soil properties using a VIS-NIR sensor. *Soil and Tillage Research* 93, 13–27.
- Mouazen, A.M., Maleki, M.R., Cockx, L., Van Meirvenne, M., Van Holm, L.H.J., Merckx, R., De Baerdemaeker, J., Ramon, H., 2009. Optimum three-point linkage set up for improving the quality of soil spectra and the accuracy of soil phosphorus measured using an on-line visible and near infrared sensor. *Soil and Tillage Research* 103, 144–152.
- Munoz, J.D., Kravchenko, A., 2011. Soil Carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs. *Geoderma* 166, 102–110.
- Nicola, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. A review—nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy. *Postharvest Biology and Technology* 46, 99–118.
- Pirie, A., Singh, B., Islam, K., 2005. Ultra-violet, visible and near-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Australian Journal of Soil Research* 43, 713–721.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* 66, 988–998.
- Shibusawa, S., Made Anom, S.W., Sato, H.P., Sasao, A., 2001. Soil mapping using the real-time soil spectrometer. In: Gerenier, G., Blackmore, S. (Eds.), *ECPA 2001*, vol. 2. Agro Montpellier, Montpellier, France, pp. 485–490.
- Shonk, J.L., Gaultney, L.D., Schulze, D.G., Scoyoc, G.E.V., 1991. Spectroscopic sensing of soil organic matter content. *Transactions of the ASAE* 34, 1978–1984.
- Stenberg, B., Viscarra Rossel, R., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy* 107, 163–215.
- Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. *Plant Soil* 251, 319–329.
- USDA Natural Resources Conservation Service (1998). *Soil Quality Information Sheet Soil Quality Indicators*: pH.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2001. Proximal sensing of soil pH and lime requirement by mid infrared diffuse reflectance spectroscopy. In: Grenier, G., Blackmore, S. (Eds.), *ECPA, Third European Conference on Precision Agriculture*, vol. 1. Agro Montpellier, Montpellier, pp. 497–508.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006a. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131, 59–75.
- Viscarra Rossel, R.V., McGlynn, R.N., McBratney, A.B., 2006b. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma* 146, 403–411.
- Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158, 46–54.
- Volkan Bilgili, A., van Es, H.M., Akbas, F., Durak, A., Hively, W.D., 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments* 74, 229–238.
- Wetterlind, J., Stenberg, B., Söderström, M., 2010. Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. *Geoderma* 156, 152–160.
- Zornoza, R.C., Guerrero, J., Mataix-Solera, S., Scow, K.M., Arcenegui, V., Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biology & Biochemistry* 40, 1923–1930.